

Psychological Monographs

General and Applied

No. 459
1958

Okarski

Consistency of Projective Movement Responses

By

Joseph F. Okarski

Teachers College, Columbia University

Price \$4.00

Vol. 72
No. 6



Edited by Norman L. Munn
Published by The American Psychological Association, Inc.

Psychological Monographs: General and Applied

Combining the *Applied Psychology Monographs* and the *Archives of Psychology*
with the *Psychological Monographs*

NORMAN L. MUNN, Editor

Department of Psychology, Bowdoin College
Brunswick, Maine

Consulting Editors

ANNE ANASTASI
FRANK A. BEACH
W. J. BROGDEN
JOHN F. DASHIELL
JAMES J. GIBSON
D. O. HEBB
EDNA HEIDEREDER
FRANCIS W. IRWIN
JAMES J. JENKINS

HAROLD E. JONES
DANIEL KATZ
BOYD McCANDLESS
DONALD W. MACKINNON
QUINN McNEMAR
HENRY W. NISSEN
LORREN A. RUGGS
CARL R. ROGERS
RICHARD L. SOLOMON

ROSS STAGGER

ARTHUR C. HOFFMAN, Managing Editor

HELEN ORR, Circulation Manager

FRANCIS H. CLARK, Editorial Assistant

Manuscripts and correspondence on editorial matters should be sent to the Editor. *Psychological Monographs* publishes comprehensive experimental investigations and programmatic studies which do not lend themselves to adequate presentation as journal articles. Major space is given to the author's original contribution; introductory and bibliographic materials, as well as statistical tables and graphs, must be kept within reasonable bounds. Tables, graphs, and appendix materials which deal with detail not essential to adequate presentation of the findings may be made available through the American Documentation Institute—for details of this procedure, see the *APA Publication Manual*. Preparation of manuscripts for publication as monographs should follow the procedure given in the *APA Publication Manual*. Publication in *Psychological Monographs* is free of cost to the author, except in cases where early publication is requested or author's alterations are made in galley proofs.

Correspondence on business matters should be addressed to the American Psychological Association, Inc., 1335 Sixteenth St. N.W., Washington 6, D.C. Address changes must arrive by the 10th of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee third-class forwarding postage.

COPYRIGHT, 1958, BY THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.

Psychological Monographs: General and Applied

Consistency of Projective Movement Responses¹

JOSEPH F. OKARSKI

Teachers College, Columbia University²

THERE has been increasing emphasis in the literature to the effect that the contradictory results encountered in validation studies of projective techniques may be due partly to the possibility that some of the underlying assumptions are sound but others faulty. For this reason it has been urged that closer scrutiny of

specific practices and their theoretical bases is needed to screen those which may be retained from those which should be modified or discarded.

Among the practices which often have been challenged has been the observation and diagnostic interpretation of movement, as conceptualized in various symbolic or representational terms furnished by language, art, and other expressive media, and converted into psychological test procedures. The Rorschach "movement response," comprising those "concepts which projected some kind of action or life into ink blots" (13, p. 109), represents perhaps the best known application of the principle.

Fundamental to the observation and interpretation of conceptualized movement in the projective techniques has been the explicit or implicit assumption that the movement response occurs as a function of something deep rooted and stable in the individual. Observed as a test response, it has thus been presumed to yield revealing and dependable clues to personality and behavior. While some evidence directly or indirectly relevant to this proposition is to be found in the experimental literature, research to date has not been conclusive. The present study, therefore, has sought to add some clarification.

Methodological Approaches

Previous approaches to the problem have usually taken one of two directions:

¹ This monograph is based on a doctoral dissertation at Columbia University. Laurence F. Shaffer, Chairman of the Doctoral Committee, and Edward V. Shoben, Jr. and Robert L. Thorndike, members of the committee, guided the development of the study as a whole. Helen M. Walker, Gordon F. Derner, and Herbert J. Zucker also contributed to the development of the study. Philip Brandis, Helen A. Diers, Joan Fox, Isidore Helfand, Shabse H. Kurland, Monroe L. Levin, Eugene Levitt, Reuben M. Margolis, Martha Hessel Page, Alfred Rubinstein, and Loron M. Simon helped in judging certain of the materials.

Special credit is due the United States Air Force which granted laboratory facilities and the cooperation of its staff under the Personnel Research Field Section Human Resources Research Center, Sampson Air Force Base, Geneva, New York, and which provided subjects for the experiment. Donald B. DuBois served as consultant on the scene.

The study is also indebted to the following: Weston A. Bousfield, who gave access to research facilities in the Department of Psychology at the University of Connecticut for important preliminary aspects of the study; The Psychological Corporation, which supplied the Levy Movement Cards; the late Oliver D. Havard, of Allentown, Pennsylvania, who gave financial support to certain phases of the study; and, finally, my wife, Ruth, who assumed many of the unenviable routine tasks and other burdens involved.

I am deeply grateful to all.

² Now at the Rochester Guidance Center, New York.

(a) attempting to establish a direct relationship between movement responses and personal attributes thought to have important and durable psychological significance and (b) examining the consistency or inconsistency of movement scores and derived indices. The search for correlates has been inconclusive partly because validity of the criteria used to define correlative variables has not always been demonstrated. More basically, however, these studies have been weak because they have uncritically assumed the movement indices provided personally relevant behavioral data largely insusceptible to chance distortion or to systematic error. Since both experimental successes and failures might conceivably be accounted for in either of these terms, it has been difficult to determine in specific instances what interpretation might be put on observed results.

Studies following the second approach have in turn been indecisive because essential standards for testing clinical hypotheses concerning stability and variation in behavior have not been met. The shortcoming becomes evident when researches are weighed against basic requirements for examining the individual's variability in relation to his presumed "true" score or typical performance.

Zubin (29) has discussed these requisites at length. Key points may be applied to the present problem as follows: (a) the individual must be regarded as an independent universe of behavior; (b) each observation (in this case, the movement score) must be treated as a random sample from this universe; (c) sampling must be sufficiently extensive to provide a definitive estimate of the individual's range of variability within his universe; (d) observations must be based on a mathematical model (e.g., ex-

pansion of the binomial) which permits any observed range of variability to be tested rigorously for its significance; (e) the performance of the individual must be treated integratively as well as differentially in relation to that of others.

Studies of Consistency in the Movement Response

Researches pertaining to intra-individual variability in the movement response have usually been Rorschach studies. Most of them have been investigations of reliability using split-half, or single-repetition test-retest designs. In some cases special operations have been introduced to test the stability of the subject's scores under varied sets, suggestions, stresses, drugs, and other experimental conditions relating either to the subject, to the examiner, or to the test situation as a whole. Miller (17) has reviewed the work in this area and has outlined various possible sources of variation in the subject's scores.

Other considerations aside: the studies of reliability have generally provided no more than two estimates or tests of the subject's performance level. Assuming that the two estimates are independent and that there is equal probability for any two alternative performance levels (e.g., high or low) in either test, then the statistical significance of complete agreement, or of complete disagreement, between the two estimates is $.5^2 + .5^2$ or only $.5$ —exactly what is expected by chance. Under these conditions it cannot be determined whether any subject has been significantly consistent or significantly inconsistent in his performance level.

Exceptions to this criticism have been longitudinal Rorschach studies like that done by Ives, Grant, and Ranzoni (11) and the study of internal consistency in

the Rorschach categories done by Wittenborn (28). Other considerations, however, have limited the implications of these studies for the present problem.

A number of researchers, for example, have suggested that, if the movement response truly involves significant personal attributes, it should reveal a broadly general character not confined to the Rorschach situation. Wittenborn (28) interpreted this hypothesis as consonant with underlying Rorschach assumptions. He suggested that, if the assumptions held, subjects might be expected to conceptualize movement with consistent frequency on a variety of tests which offered perceptual alternatives. Sanford (22) suggested that subjects who projected action into inkblots might be expected to inject action into scenes and situations which they depicted verbally. Others have likened the verb-adjective quotient used in analyzing verbal materials to the Rorschach movement-color ratio.

Going beyond the point of speculation, several investigators have offered presumably alternative or supplementary inkblot test forms, e.g., the Behn (31) and Harrower (8) series, claimed to yield results consistent with those of the Rorschach. Others have adopted or modified the Rorschach model for the experimental and diagnostic observation of movement in other types of tests involving interpretation or production of drawings, paintings, mosaic designs, toy arrangements, etc. The hypothesis of intra-individual consistency in conceptualizing movement might be extended to these techniques as well.

Several studies have offered direct evidence bearing on the consistency of the subject in movement production as observed in comparisons between different tests—the Rorschach has served as the

common denominator in these researches. The broadest was that of Mayman (15) who made comparisons between the Rorschach inkblots, on the one hand, and the Behn (31) and Harrower (8) series on the other. Extending beyond the inkblot frame of reference, King (12), and Allen, Ray, and Poole (1) compared Levy Movement card performance with that on the Rorschach. Hays, Gellerman, and Sloan (9) compared movement indices on the Rorschach with verb-adjective quotients in TAT protocols.

Since completion of the present study, Epstein, Nelson, and Tanofsky (7) have reported an exceptional investigation using ten sets of newly devised inkblots totaling 100 in all; apart from this, studies have been subject to the same weakness stressed in connection with the Rorschach studies described earlier: inadequate samplings of the individual's universe of test performance. Their results, therefore, have been equally indecisive. The question substantially remains: Does the movement response occur with such intra-individual consistency as to warrant the assumption that it represents some intrinsic, stable, personal variable? Resolution of the question would not only help to clarify the status of the movement response with respect to its basic assumptions in the Rorschach and similar tests, but might also contribute to the development of sound theory and practice in the projective techniques generally. The present study, therefore, has undertaken to explore the question according to some of the principles outlined by Zubin (29) and described earlier.

Hypothesis of the Present Experiment

It is hypothesized that, if the movement response, broadly defined, is a func-

tion of something deep rooted and stable in the individual, then his level of movement production as compared to that of other individuals is consistent over a range of different test situations.

PROCEDURES

As a general frame of reference, a movement response was defined as the verbal stipulation of some action. Subject to this limitation and to the modifications described below, conceptualized movement was observed in six different tests, all administered and scored by the experimenter. The tests were of two types conveniently designated as *Perceptual* and *Verbal*; three of each type were used.³

Perceptual Techniques

The perceptual tests included the Rorschach (20), Behn-Rorschach (31), and Levy Movement (21) cards, all administered and scored according to a standard procedure based on the Klopfer and Kelley (13) Rorschach manual. Besides the usual *M* (human), *FM* (animal), and *m* (inanimate) movement scores, a combined or sum movement (*SM*) score was obtained. All scores were converted to percentages to correct for individual differences in total number of responses per test and to provide a common base for comparison.

Verbal Techniques

The verbal techniques included two sentence construction procedures espe-

cially devised for the experiment. Stimulus words for these tests were chosen to provide cues minimally suggestive of response. It was postulated that, as the number of cues in the stimulus material was reduced, the likelihood of response to personal cues was increased. This procedure ostensibly gave greater assurance that response was more a function of the individual than of the stimulus material itself. The third verbal technique was a modification of Severn's (23) Heteronym Test. Brief descriptions follow.

Sentence Construction 1. The subject was asked to make up and write down as many short simple sentences beginning with the word *the* as he could in 30 minutes, using a "different idea" for each sentence. The test was presented as an experiment in "imagination," and the subject was encouraged to "stretch" his imagination as much as he liked and in any way he wished.

Sentence Construction 2. The subject was asked to compose one sentence for each of 40 different stimulus words like *the*, *and*, *but*, *for*, etc., incorporating these in any way he wished. Otherwise, instructions were substantially the same as those for Sentence Construction 1.

Heteronym Test. In the Heteronym Test the subject was presented visually with 40 two-syllable words, each having both verb and nonverb meanings, depending on the syllable stressed. Each was hidden in a matrix of scrambled letter mounted on a 4" × 6" card. The cards were arranged in four ordered sets including alternate cards containing nonheteronyms to mask the nature of the test. The cards were administered at intervals according to a predetermined plan (see *Administration of the Battery*). The subject was instructed to find the hidden word on each card and to say it aloud on doing so.

³ Only brief descriptions of the test procedures and materials are given here. Further details may be obtained by consulting Appendixes A-K, pp. 68-97, of the original publication: Okarski, J. F., Consistency of projective movement responses. Doctoral dissertation, Columbia University, 1956 (or Publication No. 17,072, Ann Arbor: University Microfilms).

Scoring

Sentence constructions. A scoring system similar to that used in the perceptual tests was experimentally developed. The method was based on adaptations of Rorschach scoring rules. A list of over 300 sample sentence constructions served as scoring criteria, e.g.: *M*—"The player made a home run." *FM*—"A cat just came in." *m*—"The tree was struck by lightning."

All sentence constructions were scored accordingly; raw scores were converted to percentages as in the perceptual tests.

The reliability of the scoring system was examined by submitting the responses of 64 subjects on Sentence Construction 1 to two independent judges, one scoring the odd numbered and one the even numbered items. Their scores were then compared with those of the experimenter. Ninety-one percent of the odd numbered items were scored with perfect agreement; 95% of the even numbered items were scored with perfect agreement.

Correlation coefficients comparing the scores given by the judges with those given by the experimenter were also computed. Corrected by Spearman-Brown formula, the coefficients ranged from .91 to .98. On the basis of these figures, the reliability of the evolved system of scoring appeared to be adequate for experimental purposes.

Heteronym Test. For the Heteronym Test, movement was scored in terms of the subject's pronunciation of the heteronym: if he pronounced the word according to the verb definition, he was credited with a movement response. The proportionate number of such pronunciations was recorded as the subject's movement score.

There appeared to be no basis for differentiating various subcategories of

movement as in the other tests; hence only the one index of movement was obtained. It was considered simply as a general indicator of movement response tendency corresponding most nearly, perhaps, with *SM* but suitable for comparison with any other type of score.

Since the technique was virtually self-scoring and quite objective, no test of scoring reliability was considered necessary.

Subjects

Complete data were obtained for 82 male subjects, 18 to 25 years of age. They were drawn from an experimental pool of military recruits in their sixth day of induction processing at Sampson Air Force Base, Geneva, New York. Subjects were required to be at least high school graduates; except for this limitation, selection was by random sampling procedures employed by the base for its own research purposes.

Subjects came from 17 different states throughout the eastern United States. Six were Negroes; the rest, white. Thirteen had up to two and one-half years of college training; 10 were college graduates. Only one reported having courses in psychology; these were of a rudimentary nature. No data concerning intelligence or other diagnostic attributes were obtained. Some notion of the social backgrounds of the subjects, however, might be inferred from the following information about parent occupation: 40% were skilled or semiskilled tradesmen or laborers in industry, business, mining, or construction; 23% were unskilled laborers; 18% held supervisory or managerial positions; 9% operated their own minor businesses; 5% were professional people; and 5% were farmers. Outside the background information given, subjects remained anonymous.

Administration of the Battery

Subjects were scheduled in pairs to complete the tests in one day. All tests except the Sentence Constructions were administered in individual sessions; the latter were given in joint sessions. Joint

TABLE 1
ODD-EVEN RELIABILITY COEFFICIENTS
FOR ALL SIX TESTS
(Based on percentage movement scores
for 82 Ss)

Test	Movement Category			
	<i>M</i>	<i>FM</i>	<i>m</i>	<i>SM</i>
Perceptual				
Rorschach	.34	.56	.25	.69
Behn	.41	.63	.42	.69
Levy	.59	.36	.49	.53
Verbal				
Sent. Constr. 1	.88	.70	.82	.87
Sent. Constr. 2	.75	.43	-.02	.74
Heteronym	—	—	—	.73

Note. — Coefficients corrected by Spearman-Brown formula.

and individual administrations were alternated to permit the subject suitable rest periods.

After a standardized orientation given to facilitate rapport and to dispel any examination set acquired by the subjects during the preceding days of classification and aptitude testing, the tests were given in the following order: (a) Heteronyms, first set; Rorschach; Heteronyms, second set; (b) Sentence Construction 1; (c) Heteronyms, third set; Behn-Rorschach; Heteronyms, fourth set; (d) Sentence Construction 2; (e) Levy Movement Cards. Before taking the tests subjects were given the option of withdrawing from the experiment without penalty in order to put participation on a voluntary basis. Brief preliminary interviews served to obtain personal data, as reported above, just before administration of the first test procedure.

Reliability of the Tests

Percentage scores on odd-even split-halves of each test were correlated with each other to obtain estimates of test reliability. In the case of the perceptual tests, scores were based on odd and even card groupings. In the verbal tests, scores were based on odd and even items.

Table 1 presents the values obtained, as corrected by Spearman-Brown formula. Coefficients varied widely, ranging from $-.02$ to $.88$. In general, the verbal tests appeared to be more reliable than

the perceptual. Sentence Construction 1 was the most reliable single test with values ranging from $.70$ to $.88$. The Rorschach was least reliable, with r running from $.25$ to $.69$.

Of the several movement indices, *SM* appeared to be the most reliable, r ranging from $.53$ to $.87$. Least reliable of the movement indices was *m*, with r running from $-.02$ to $.49$, excepting the single value, $.82$, obtained for Sentence Construction 1.

On the basis of these estimates, the techniques chosen for the experiment did not, as a whole, appear to be highly reliable. As reliability coefficients, however, the estimates presumed the measurement of a psychological variable which was relatively stable in character. It was precisely this feature, the stability of performance, which the experiment brought into question. It proposed that the answer to the question might be determined by observing the generality of the subject's performance among successive samples of behavior, as defined by each of several tests. The split-halves of each test, however, might also be regarded as successive samples or subsamples of behavior among which the generality of the subject's performance might be observed. The generality of the subject's performance within the limits given by split-half comparisons might thus be equally regarded as evidence concerning the relative stability or instability of the variable *within* each test.

Treatment of the Data

It was assumed that a particular form of behavior represents something deep rooted and stable in the individual if it occurs repeatedly in different situations with such regularity as to minimize chance occurrence. Treatment of the data, therefore, hinged on some method

of classification which permitted probability calculation.

Classification of scores and probability. All percentage scores were first classified as high or low, depending on whether they fell above or below their respective medians as shown in Table 2. Scores falling close to the median were computed to three or more decimal places to eliminate ties and determine classification. Where this procedure failed to break the distribution into equal halves, near-median scores were assigned at random to equalize the two class frequencies. Tied scores falling at the median in the larger of the approximate halves were serialized, and the number needed to fill out the smaller halves were chosen through a table of random numbers. The procedure was followed for all distributions otherwise failing of equal halves, except where the median fell at zero—in these cases, all zero scores were classified as low, and the rest as high.

Of the 21 distributions shown in Table 2, 12 required no random allocations. In the remaining distributions, the number of scores allocated at random from the larger half ranged from 1 to 4, with the average at 1.9.

Except for the *FM* category in the Levy test and Sentence Construction 2, where more than half of the subjects received zero scores, the classification procedure divided the sample in each movement category into two equal groups of high and low scoring subjects and defined equal probability (.5) for high and low scores in each test. In the *FM* category for the two tests just mentioned, proportions of high and low scoring subjects were .37 + .63 and .22 + .78 respectively, defining corresponding probabilities. Assuming each test to be an independent sampling of behavior, the prob-

ability of any high-low score pattern or combination in any category of movement could thus be computed either by expanding the binomial or by enumerating all combinations and permutations of high and low proportions. The procedure made simultaneously available (a) tests of significance for any amount of variation or consistency between high and low scores as shown by any one subject and (b) chance expectancy values against which the performance of the group as a whole could be evaluated.

Treatment of the data on the basis described involved certain dangers which should be noted. On the one hand, dichotomizing scores at the median sometimes defined wide limits within which the subject's scores might fluctuate and still be regarded as consistent as long as they were either above or below the median. On the other hand, if the subject's scores fluctuated about the median, they were automatically defined as inconsistent even though they might fluctuate within a narrower range than scores which otherwise qualified as consistent. The basis for treatment of the data thus raised the distinct possibility that consistent performance might not be detected in some cases and that individual differences in variation, as defined, might frequently represent artificial distinctions rather than true differences.

The method thus incurred serious risks of error. The importance of the experimental problem, however, appeared to justify some exploratory risks. In addition, in view of the way in which the data were distributed, it was questionable whether any other method was more appropriate. It was decided, therefore, to accept the risks involved while emphasizing that due caution might be necessary in interpreting the results.

TABLE 2
DISTRIBUTION OF ALL PERCENT MOVEMENT SCORES
(R: Rorschach; BR: Behn-Rorschach; LM: Levy Movement; S: Sentence Construction;
H: Heteronyms; $N=82$)

Score	M					FM				
	R	BR	LM	S ₁	S ₂	R	BR	LM	S ₁	S ₂
.99-.100										
.96-.98										
.93-.95										
.90-.92										
.87-.89										
.84-.86										
.81-.83										
.78-.80										
.75-.77			2							
.72-.74			0							
.69-.71			1							
.66-.68			1							
.63-.65			1							
.60-.62			0		2					
.57-.59			1		3					
.54-.56			1		3		2			
.51-.53			0		3		0			
.48-.50			4		4		1			
.45-.47			2	1	6		1			
.42-.44			2	1	6	1	1			
.39-.41			3	0	6	2	2			
.36-.38			3	0	6	3	5			
.33-.35	2		4	1	12	1	3			
.30-.32	0	1	2	0	1	4	5			
.27-.29	1	1	5	1	4	5	6			
.24-.26	2	0	6	1	10	12	7			
.21-.23	4	0	4	3	6	8	5			
.18-.20	8	7	12	5	7	11	12	2	1	
.15-.17	9	8	1	8	2	7	0	0	0	
.12-.14	10	11	9	5	1	4	8	5	1	
.09-.11	7	7	7	13		8	3	4	2	
.06-.08	16	10	3	21		3	4	9	4	1
.03-.05	8	12	1	14		6	3	10	38	17
.00-.02	15	16	7	8		7	5	52	36	64
\bar{X}	.11	.00	.25	.10	.35	.18	.22	.03	.03	.01
Mdn	.00	.07	.21	.08	.35	.10	.10	.00	.02	.00
S	.08	.07	.18	.09	.12	.11	.12	.05	.03	.02

Plan of comparison. If the experimental hypothesis were generally true, it would be expected that for any category of movement all subjects, or a significantly large majority, would score *either* high *or* low on all six tests with a regularity beyond that anticipated on the basis of chance.

If the hypothesis were not generally true, then it would be expected that subjects would shift between the two performance levels, with high-low score combinations showing either uniformly

wide variation or individual differences distributed substantially according to chance.

To determine the status of the hypothesis, therefore, the high-low score patterns of all subjects were classified and proportionately distributed in each category of movement according to combinatorial order and corresponding probability. Chi-square tests (27) were then applied in each category to determine the nature of each distribution in relation to chance. The .05 confidence

TABLE 2 (continued)

Score	MI					SM					
	R	BR	LM	S ₁	S ₂	R	BR	LM	S ₁	S ₂	H
.99-.100								1			
.96-.98								0			
.93-.95								0			
.90-.92								0			
.87-.89								1			
.84-.86								0			1
.81-.83								0			0
.78-.80								0			2
.75-.77						1		1			2
.72-.74						0		0			3
.69-.71						0		2			6
.66-.68						1		2			6
.63-.65						1	2	0		1	12
.60-.62						1	2	2		3	2
.57-.59						2	1	1	1	5	5
.54-.56						1	3	1	1	3	6
.51-.53						2	4	1	1	4	11
.48-.50						3	5	6	2	6	8
.45-.47						4	4	4	0	6	4
.42-.44						5	7	4	4	6	6
.39-.41						5	5	4	5	3	3
.36-.38				1		7	2	0	3	0	2
.33-.35				0		8	8	8	4	7	2
.30-.32				1		8	11	4	5	8	1
.27-.29				0		3	6	8	5	4	
.24-.26			3	2		4	6	8	10	8	
.21-.23			1	5		7	1	4	12	3	
.18-.20		1	5	3		6	1	6	6	4	
.15-.17		0	4	10		4	2	1	6	1	
.12-.14	2	1	9	10		1	4	2	7	1	
.09-.11	5	5	9	10		1	1	1	4		
.06-.08	12	17	8	10	4	2	3	0	3		
.03-.05	21	23	8	6	39	2	1	0	2		
.00-.02	42	35	35	6	39	3	3	4	1		
\bar{X}	.03	.04	.07	.12	.02	.32	.34	.36	.25	.38	.57
<i>Mdn</i>	.02	.03	.06	.10	.02	.32	.33	.33	.23	.38	.55
<i>S</i>	.04	.04	.08	.07	.02	.16	.16	.18	.12	.12	.12

level was used to define the boundary between a chance and a nonchance distribution.

Supplementary analyses were undertaken to detect any significant trends and relationships in performance over the three verbal tests, the three perceptual tests, and all possible pairs of tests.

Finally, the relation of the subject's variability in one category of movement to that in others was examined.

Subject sampling. Comparisons were made for two alternate groups of sub-

jects. The first comprised all 82 subjects as described earlier. The second comprised a subsample of 64 of these same subjects who gave a minimum of 15 Rorschach or 15 Behn-Rorschach responses. The purpose of the alternative sampling was to observe the effect of error due to the use of percentage movement scores based on small response totals.

RESULTS

It is hypothesized that, if the conception of movement is a function of some-

TABLE 3

SIX-TEST PERFORMANCE: PROPORTIONATE DISTRIBUTION OF HIGH-LOW MOVEMENT SCORES OBSERVED AND EXPECTED BY CHANCE IN EACH MOVEMENT CATEGORY FOR ALL SUBJECTS
($N=82$)

Movement	Proportionate frequencies in each category of movement					
	Probable		Observed			
H-L	M , m , SM	FM	M	FM	m	SM
6-0	.016	.005	.024	.000	.037	.037
5-1	.004	.046	.122	.061	.134	.122
4-2	.234	.167	.256	.207	.195	.207
3-3	.313	.303	.232	.280	.280	.256
2-4	.234	.297	.195	.244	.183	.232
1-5	.004	.150	.134	.171	.122	.110
0-6	.016	.031	.037	.037	.040	.037
χ^2 Sig. ^a			7.25 .50	2.84 .90	11.94 .10	6.63 .50

^a Six degrees of freedom.

thing deep rooted and stable in the individual, then his level of movement production, as compared with that of other individuals, is consistent over a range of different test situations. Analysis of the data gave no support to the hypothesis as a general proposition, as shown by the results below.

Six-Test Performance

Table 3 shows the results for all 82 subjects on the full test battery. Given is the proportionate distribution of high-low score patterns probable and observed in each movement category, with the derived chi square and significance values. Individual patterns varied from constantly high (6-0) to constantly low (0-6). Only at these extremes could the subject's performance be regarded as significantly consistent at the .05 level or better. Less than 4% of the subjects in any movement category scored constantly high; less than 5% in any instance scored constantly low. Distribution chi squares

TABLE 4

SIX-TEST PERFORMANCE: PROPORTIONATE DISTRIBUTION OF HIGH-LOW MOVEMENT SCORES OBSERVED AND EXPECTED BY CHANCE IN EACH MOVEMENT CATEGORY FOR THE SUBSAMPLE
($N=64$)

Movement	Proportionate frequencies in each category of movement					
	Probable		Observed			
H-L	M , m , SM	FM	M	FM	m	SM
6-0	.016	.007	.031	.000	.047	.047
5-1	.004	.057	.094	.078	.150	.125
4-2	.234	.185	.297	.250	.203	.203
3-3	.313	.309	.250	.266	.297	.266
2-4	.234	.282	.172	.203	.156	.219
1-5	.004	.135	.125	.172	.109	.109
0-6	.016	.026	.031	.031	.031	.031
χ^2 Sig. ^a			5.60 .50	4.98 .75	9.82 .25	6.62 .50

^a Six degrees of freedom.

were not significant at the .05 level, indicating no great discrepancy between observed and chance frequencies.

Table 4 shows a corresponding analysis for the subsample of 64 subjects whose Rorschach and Behn-Rorschach records showed response totals of 15 or more. High-low score patterns were distributed in substantially the same way as those for the larger sample. Distribution chi squares were again insignificant at the .05 level.

Both samplings thus showed chance distribution of individual differences in variation between high and low movement scores. In these terms, it was apparent that, if the hypothesis held, it did so for a few cases only and not for the great majority of subjects. Furthermore, since frequencies were distributed generally according to chance, there was no assurance that even these few cases were not accidental.

It was also apparent from the two subject samplings that the results of the an-

alysis were not seriously affected by either the inclusion or exclusion of subjects with low response totals. Succeeding analyses revealed similarly duplicating results for the two samples; for this reason, and because the samples were not otherwise independent, only the results for the larger group are hereafter presented.

To determine whether the few cases who seemed to be consistent represented the hypothesized phenomena, the data for the six tests were analyzed further to detect any supportive trends. For this purpose the high-low score combinations of all subjects were redistributed according to the degree of variation or maximum intertest agreement and disagreement represented. In effect, this procedure merely served to combine distribution frequencies for corresponding extreme intervals as shown in Table 3. Thus, proportionate frequencies for intervals 6-0 and 0-6 were combined into a single interval representing six tests in agreement; frequencies for intervals 5-1 and 1-5 were combined in a single interval representing five tests in agreement, etc. Chi-square tests were then applied as before.

Only in the *m* category was a trend in the hypothesized direction observed. Roughly 9% of the subjects, as compared to an expected 3%, were significantly consistent with all six tests in agreement on movement score level. About 26%, as compared to an expected 19%, showed five tests in agreement. The remaining proportions fell below the expected values. The distribution chi square was 11.46, significant at the .001 level. Thus, there was no support for the hypothesis except in *m* production.

For a still coarser analysis the number of subjects in each movement category who were significantly consistent (six

tests in agreement) were compared with the number who were not significantly consistent (three to five tests in agreement). This involved combining proportionate frequencies for intervals 6-0 and 0-6, as shown in Table 3, in one cell and combining frequencies for the remaining intervals in a second cell. Chi square was then computed to determine whether the observed number of significantly consistent subjects exceeded chance expectancy. Again only in *m* were the results significant. Chi square with Yates' correction was 6.2 with a probability of .02.

On the basis of the coarser analysis, the hypothesis appeared to be confirmed for a small number of subjects on *m* production, and was rejected for the rest. It might be concluded that the conception of movement can be safely assumed to represent some intrinsic, stable variable for a small number of subjects only, and then only with respect to inanimate movement responses.

Supplementary Analyses

It was apparent from the previous results that the subject's movement score generally varied from test to test. A question thus arose as to the roles played by specific tests or types of tests in determining the outcome of the experiment. Analysis was therefore narrowed to the areas of performance defined by the three verbal tests, the three perceptual tests, and all possible pairs of tests.

Three-Test Performance

The high-low score patterns for each set of three tests were examined exactly as were the data for the six tests. Only for the perceptual tests were the results significant, as shown in Table 5 which gives the proportionate high-low score distribution in each movement category.

About 17% to 22% of the subjects

scored constantly high; about 20% to 23% scored constantly low. All these proportions were greater than chance values, while those for intervening intervals were lower. Distribution chi squares were all significant at the .01 level or better. It was thus apparent that, for every movement category, significantly large numbers of subjects were consistent in level of movement production on the perceptual tests.

Combining distribution values according to the maximum intertest agreement represented by high-low score combinations, it may be seen that the probability for consistent performance in any one movement category was .25. In other words, perceptual test samplings did not provide a crucial test as to the significance of any one subject's performance. Hence, while significantly large numbers of subjects were consistent in these tests, it could not be said that any subject was significantly consistent at the standard .05 confidence level.

Trends in the hypothesized direction were nevertheless evident for every movement category. In this respect, the results for the perceptual tests contrasted with those for the verbal tests. It might be concluded that performance in these tests gave more support to the hypothesis of consistency in movement production and that the conception of movement was more likely to be a function of something deep rooted and stable in tests of this character, perhaps, than in those of the verbal type. It appeared unlikely, however, even on these terms, that the hypothesis could hold equally for all or even a majority of subjects.

Paired-Test Comparisons

Using four-fold contingency tables, phi coefficients and chi squares were

TABLE 5
PERCEPTUAL TESTS: PROPORTIONATE DISTRIBUTION OF HIGH-LOW MOVEMENT SCORES OBSERVED AND EXPECTED BY CHANCE IN EACH MOVEMENT CATEGORY
($N = 82$)

Movement	Proportionate frequencies in each category of movement					
	Probable		Observed			
H-L	<i>M</i> , <i>m</i> , <i>SM</i>	<i>FM</i>	<i>M</i>	<i>FM</i>	<i>m</i>	<i>SM</i>
3-0	.125	.091	.220	.183	.171	.220
2-1	.375	.341	.256	.232	.390	.280
1-2	.375	.409	.320	.354	.207	.280
0-3	.125	.159	.195	.232	.232	.220
χ^2			12.63	13.83	15.04	15.63
Sig. ^a			.01	.005	.005	.005

^a Three degrees of freedom.

computed to determine the interrelationship of high and low scores in each category of movement for all possible pairs of tests. Table 6 shows the results. Coefficients ran uniformly low, ranging from .41 to -.27 with 19 out of 60, or nearly a third, on the negative side. All together, only 9 showed significant interrelationships in performance.

Significant phi values were largely confined to the perceptual tests, as might have been expected from the results for three-test comparisons. Performance on the Rorschach was significantly related to that on the Behn-Rorschach in every movement category and to that on the Levy Movement Cards in the categories *M* and *m*. In contrast, there were no significant relationships between the Levy and the Behn, though all phi values were positive. There were no significant relationships between any two verbal tests and few for verbal-perceptual pairings.

It is notable that, while the only significant result observed in six-test performance was in *m*, paired-test compari-

sons showed most of the nine significant phi values in *M*. Included were the only significant relationships observed in verbal-perceptual comparisons; one of these, however, was significantly negative: $-.27$ for the Levy vs. Sentence Construction 2.

It may be concluded from these and from the foregoing results that, if the conception of movement bears any significant relationship to psychological processes in the individual for more than a minority of subjects, then the relationship is largely specific to certain tests or types of tests and that performance is not generally predictive from one type of test to another for any individual.

Intercategory Comparisons

Results of the experiment showed the presence of individual differences in variation between high and low movement scores. These differences were so distributed that the hypothesis was, in general, only tenuously upheld for some subjects and not for others *within* each movement category. Two further questions thus suggested themselves: (a) If movement scores were not themselves generally stable so as to justify the assumption that something deep rooted and stable was reflected, was there anything to suggest that something of an intrinsically stable character was represented by individual differences in variation? (b) Since the experimental hypothesis was upheld (even though tenuously) for some subjects in each movement category, was there anything to suggest selective consistency with possibly stable individual differences generally defined on the basis of the specific category or categories in terms of which consistent tendencies were selectively expressed?

Comparisons between the several

movement categories were undertaken in the hope that some answers to these questions might be obtained.

Individual Differences in Variation

It is hypothesized that, if individual differences in variation between high and low movement scores represent intrinsically stable characteristics, then the degree of variation shown by the subject in one category of movement is consistent with that shown in other categories of movement.

To examine the hypothesis, the data were first treated as in paired-test comparisons, except that movement categories instead of tests were paired, and corresponding individual differences in variation, instead of classified movement

TABLE 6

PAIRED TESTS: PHI VALUES FOR PAIRED-TEST COMPARISONS

(R: Rorschach; BR: Behn-Rorschach; LM: Levy Movement Cards; S₁: Sentence Construction 1; S₂: Sentence Construction 2; H: Heteronyms; *N* = 82)

Paired tests	Movement category			
	<i>M</i>	<i>FM</i>	<i>m</i>	<i>SM</i>
Verbal				
S ₁ -S ₂	-.02	.00	.12	.07
S ₁ -H	-.02	-.12	.02	-.07
S ₂ -H	-.12	.00	-.07	-.17
Perceptual				
R-BR	.31*	.31*	.22*	.41*
R-LM	.22*	.15	.22*	.17
BR-LM	.12	.20	.17	.17
Verbal-Perceptual				
S ₁ -R	.02	.02	.12	.07
S ₁ -BR	.31*	.07	.12	.17
S ₁ -LM	.07	-.15	-.02	-.02
S ₂ -R	-.07	-.06	.17	-.12
S ₂ -BR	.02	-.12	.07	-.17
S ₂ -LM	-.27*	-.10	.07	.17
H-R	.17	.12	.02	.17
H-BR	.27*	-.12	.02	.07
H-LM	.02	-.06	.07	.12

* Significant at the .05 level as indicated by chi square with one degree of freedom.

TABLE 7
INTERCATEGORY COMPARISONS: PROPORTIONATE
DISTRIBUTION OF SUBJECTS, PROBABLE AND
OBSERVED, ACCORDING TO THE NUMBER OF
MOVEMENT CATEGORIES—*M*, *FM*, OR *m*—
IN WHICH THE SUBJECT WAS RELATIVELY
CONSISTENT ON THE FULL BAT-
TERY OF TESTS
(*N* = 82)

No. of categories	Proportionate frequencies	
	Probable	Observed
3	.020	.073
2	.198	.122
1	.444	.451
0	.328	.353
χ^2		8.03
Sig.		.05 ^a

^a Three degrees of freedom.

^b One degree of freedom.

scores, were plotted against each other in contingency tables. Individual differences in variation were defined by the maximum interest agreement shown in high-low score combinations. Chi squares were then computed to determine whether degree of variation in one category of movement was significantly related to that in others, for the verbal tests, the perceptual tests, and the full battery.

Chi squares were significant for *M* to *SM* comparisons only. Excepting these two categories, therefore, there was no indication that individual differences in variation were significantly related from category to category. The hypothesis was thus rejected as a general rule. Inspection of raw score data indicated that the connection between *M* and *SM* indices was largely a function of part-whole score relationship and so offered no exception to this conclusion.

The nature and effect of the part-whole score relationship was most clearly evident in the Heteronyms, Sentence Construction 2, and Levy Movement

tests. In the Heteronym Test, *FM* and *m* responses were not differentiated; only the general, *SM* score was obtained. Since the *SM* score was entered into all comparisons, the subject's *M* and *SM* scores were necessarily identical. Similar situations obtained for the Sentence Construction 2 and Levy tests because relatively few *FM* and *m* responses were evoked. Since the subject's *M* and *SM* scores were thus identical—or nearly so—in three of the six tests, individual differences in variation for both movement categories were heavily biased toward duplication. The significance of the relationship between them, therefore, was more indicative of limitations in the test procedure than of intrinsically stable characteristics in the individual.

Extending the analysis, all subjects were classified and distributed according to the number of categories in which movement production level was relatively consistent. Since *SM* was not independent of all three common categories—*M*, *FM*, and *m*—it was excluded, which allowed a range of distribution from 0 to 3. For the full battery, performance was considered relatively consistent if five to six tests were in agreement; for both verbal and perceptual techniques, agreement among all three tests was required to meet the criterion. Exact probabilities were computed for each distribution interval by combining and permuting proportions observed to be relatively consistent, as defined above, within each category. A chi-square test of each distribution was then made to detect any significant discrepancy between probable and observed frequencies.

Only the full battery revealed any significant results. These are shown in Table 7, giving probable and observed distributions for both grouped and un-

grouped intervals. Observing the ungrouped data, it may be seen that about 7% of the subjects were, in fact, relatively consistent in all three movement categories, in comparison to a probable proportion of .029. The data thus showed that a limited number of subjects were, in fact, generally consistent in level of movement production beyond the likelihood of chance-determined performance at better than the .05 level of confidence.

The distribution chi square was 8.03, significant at the .05 level. Since the remaining observed proportions were relatively close to or below the expected values, it might be inferred that the number of subjects who were generally consistent was significantly greater than zero, even though small in absolute terms.

No other significant inferences could be drawn from these results. The hypothesis was thus tenable on a limited basis only. It might be concluded that individual differences in variation represented something of a stable character only for a small but significant number of subjects who were generally consistent in level of movement production. It might also be asserted that on the basis of these intercategory comparisons the limited and tenuous support found earlier for the original experimental hypothesis was confirmed and strengthened, and that the conception of movement in all three common categories, *M*, *FM*, and *m*, might well represent something deep rooted and stable for certain extreme cases, though not for the great majority of subjects.

Selective Consistency

It is hypothesized that if deep rooted and stable individual differences are ex-

pressed in terms of selective consistency, then all or a significantly large majority of subjects are relatively consistent in some one or more of the several movement categories.

It may be seen from the grouped data in Table 7 that a majority, or about 65% of the subjects, were relatively consistent in 1 to 3 movement categories; 35% were consistent in none. Chi square was .23, indicating that these proportions were almost exactly what was expected by chance.

There was thus no support for the hypothesis, even though a majority of subjects were consistent in one or more movement categories. Results obtained separately for the verbal and perceptual tests, not shown here, failed similarly to support the hypothesis. It was concluded that neither for the verbal, the perceptual, nor for the full battery of tests was it possible that intrinsically stable individual differences were defined in terms of selectively consistent tendencies favoring some one or more movement categories.

Summary of Results

The over-all performance of the subjects on the full test battery failed to support the hypothesis of consistency as a general proposition. Analysis showed the presence of widely distributed individual differences in variation between high and low movement scores for every category of movement. In these terms, the hypothesis was tenuously upheld, in most movement categories, for some few subjects only.

Some exception was apparent in *m* production where a small but significant number of subjects were significantly consistent. It was concluded that the conception of movement might safely be

assumed to represent something deep rooted and stable for a small number of individuals only and then only with respect to *m* production.

Intercategory comparisons of individual differences in variation, however, revealed a small but significant number of subjects who were generally consistent in all three common categories, *M*, *FM*, and *m*. On this basis, the above conclusion was somewhat strengthened and broadened to suggest that the conception of movement in all three categories might well represent something deep rooted and stable for certain extreme cases, though not for the great majority of subjects. Otherwise, individual differences in variation gave no evidence of intercategory stability so as to suggest that these differences might define something of a psychologically significant, stable character in the individual. There was also no evidence that stable individual differences might be defined in terms of selectively consistent tendencies favoring some one or more movement categories.

When trends and relationships among specific tests or types of tests were considered, positive results were confined almost exclusively to the perceptual tests. These showed trends in the hypothesized direction for every movement category. It was concluded that, if the conception of movement was a function of something deep rooted and stable in the individual, then it was more likely to be observed in tests of the perceptual than of the verbal type; but even on this basis, it could not be assumed to hold equally for all or even a majority of subjects.

Paired-test comparisons showed that trends among the perceptual tests were largely confined to significant relation-

ships between the Rorschach and Behn-Rorschach. All paired-test relationships, however—perceptual, verbal, and verbal-perceptual—were of a low order. It appeared that, if the conception of movement was related to something deep rooted and stable for more than a minority of subjects, then the relationship was largely specific to certain tests or types of tests and that performance could not be regarded generally predictive from one type of test to another for any individual.

DISCUSSION

Is the movement response generally stable? The answers suggested by previous studies might be summarized as follows: that the movement response is (a) generally stable, (b) generally unstable, (c) stable under certain conditions and unstable under certain conditions, (d) stable for some individuals or groups but not for others.

Under the more or less varied conditions represented by the different tests used in the present study, the experimental results might be said to agree that the movement response is *not* generally stable. They appeared to suggest, further, that the status of the movement response may vary depending on factors inherent in both subject sampling and the conditions under which the subject's movement responses are observed. The results might best be discussed, therefore, from the standpoints of the basic components of each of these sources of variation: (a) the individual and (b) the tests or types of tests used to sample his universe of movement responses. Some further implications may then be drawn in relation to general theory and practice in the projective techniques.

The Movement Response in Relation to the Individual

The results of the experiment suggested that when the individual is regarded as an independent universe of behavior and suitable tests of significance are applied to his performance over a range of different test situations, the conception of movement may well represent a significantly consistent variable for certain subjects. In these terms, the experimental results gave partial support to the predictions expressed by Wittenborn (28), Sanford (22), and others and encouraged the belief that the movement response as observed under a variety of different conditions or tests may well represent an intrinsically stable personal variable for some individuals or groups.

All three movement response categories—*M*, *FM*, and *m*—appeared to have personal relevance on these limited terms. The weight of evidence, however, was such as to emphasize most strongly that none of the commonly observed movement categories may be safely assumed stable and, hence, personally relevant for all individuals or groups in any or all test situations, including tests such as the Rorschach.

Question might be raised, however, as to whether the implication noted was not purely a function of the experimental method used. It may be recalled that a weakness of the method, as discussed earlier, was the possibility that consistently performing subjects might fail to be detected if their scores fluctuated about the median. It is entirely possible that in confirming the hypothesis of consistency for some subjects and not for others the experimental method merely pointed out those whose scores were, on the average, either more or less extreme

rather than more or less consistent. Due caution, thus, urged verification of the finding by further research.

While the experimental results did not foreclose the possibility that the movement response might be generally stable, they strongly discouraged such a possibility. In so doing, they followed a pattern common to a number of studies in which some effort—formal or informal—was made to examine either the stability or the validity of the movement responses and/or derived indices on the basis of the individual as well as group statistics. Milton (18), for example, found that both the *M* scores and the *M:C* ratios on the Rorschach were stable for only about a third of his subjects whether retested under experimentally induced set or under standard conditions. Using backward and forward test-retest Rorschach administration, Rabin and Sanderson (19) found the *M:C* ratios stable for 80% of their subjects and unstable for the rest. Thornton and Guilford (25) found split-half agreement for *M:C* for only 55% of their subjects. Altus and Thompson (3) used a standard administration of the Rorschach and found that, while consistent relationship between *M* scores and a test of intelligence held "almost invariably" for subjects whose movement scores were high, it failed to hold generally for those whose scores were low. The results of the present study, together with some of the more careful studies reported in the literature, thus strongly suggested that the movement responses and derived indices may be personally relevant for some individuals or groups but not for others and, further, that this conclusion may hold regardless of whether the movement responses are observed on an intra- or intertest basis and whether constant

or varied conditions are maintained.

It might also be asserted on the basis of the studies cited that, in confirming the hypothesis of consistency for some subjects but not for others, the general pattern of results yielded by the present study was not unique to its experimental approach. Thus, while due caution on the grounds described earlier strongly urged further verification, there appeared to be no reason why the findings should not be tentatively accepted.

Experimental and practical implications. If the findings are indeed correct, they may help to understand the conflicting and otherwise inconclusive implications generally reported on the basis of group results alone. For if the movement responses are stable or relevant for some individuals or groups but not for others, then, apart from any other sources of error, the positive or negative results observed in a given study may simply reflect the vagaries of unwittingly selective subject sampling. Considered in this light, the widely varied estimates sometimes reported for both the reliability and validity of the movement responses may not reflect as much on the doubtful nature of the evidence as on its sampling limitations.

The crucial question with respect to the reported findings may thus be not so much *whether* the movement responses provide dependable indices to personality and behavior but, rather, *for whom?* Practical considerations would thus seem to demand re-investigation of the findings under individual-oriented procedures to identify, if possible, those types of individuals or groups for whom the movement responses may safely be assumed relevant and to provide a discretionary basis for clinical interpretation.

The present study offered no basis for

testing any interpretive hypothesis such as might prove valid within the limits suggested. Those hypotheses in terms of which some significantly positive results have been reported, however, might be considered plausible: Altus and Thompson (3), Tucker (26), and others have found significant but low relationship between Rorschach movement indices and intelligence. Singer, Meltzoff, and Goldman (24), Meltzoff and Levine (16), and others found some significant relationship between Rorschach movement responses and motor inhibition, inner control, and fantasy. Altus (2) and others have reported positive relationship between Rorschach movement indices and evidences of maladjustment and psychopathology.

It remains to be established, however, for what classes of subjects these studies may offer dependable clinical hypotheses. From the nature of its subject sampling and the results of the present study, it might only be inferred that the movement responses given by subjects from some ostensibly normal or nonpsychiatric groups may rarely be found sufficiently consistent to warrant any kind of interpretation. It might be inferred further that, pending identification of the sampling characteristics or other convenient discriminative criteria which might define the psychologically significant populations, the clinician may have no assurance that any interpretation offered in a particular case actually relates to anything typical of the subject.

Indices of variation as diagnostic variables. A number of investigators have suggested that not only the consistency but the inconsistency and perhaps individual differences in amount of variation observed in clinical or test behavior may serve to identify or discriminate signifi-

cant classes of individuals and personality variables. Thus Hutt, Gibby, Milton, and Pottharst (10) and others suggested that the relative consistency or inconsistency of the Rorschach scores might differentiate, respectively, the "rigidity" of the abnormal subject from the "capacity to shift" or flexibility of the normal. Dörken (5) suggested modifications on a similar hypothesis depending on whether individual differences in variation were observed on an intra- or intertest basis involving projective techniques of different types.

The crucial question with regard to any proposals of this character is whether or not individual differences in variation are themselves generally stable and thus define true differences and not merely chance distinctions. Reviews of the literature on projective techniques reveal no effort to deal with this important question. The present study offered implications relating to the problem in terms of intercategory comparisons of individual differences in variation between high and low levels of movement production. The danger of using this method of observing individual differences in variation has already been noted. In the absence of any previous effort to deal with the problem of their stability in the projective techniques, however, the tentative results may be of considerable interest.

The results suggested that individual differences in variation were not significantly related from one area of performance to another, and that there was no basis for assuming that the differences were of general psychological import or that they defined a psychologically significant class of subjects or personality variables in any instance—except where the subject's performance was generally

consistent. The results thus offered no support among the movement indices for any proposals of the character suggested. At the same time, the results warned against making any diagnostic inferences on the basis of individual differences in variation without first checking on the stability of those differences.

The data did not preclude the possibility that individual differences in variation might be of a stable character *within* specified areas of performance as defined, for example, by each of the several categories of movement examined in the present study. It is not inconceivable that interpretive inferences might be warranted on such a basis. Confirmation of this possibility in the present experiment, however, would have required replication of the experiment with the same subjects and was beyond its scope. The possibility thus remains for further investigation.

The Movement Response in Relation to the Test or Type of Test

While the over-all results of the experiment suggested that interpretation of the movement responses given by ostensibly normal subjects may be justified for some few cases only, subsidiary analyses suggested that the number might be expected to vary depending on the test or type of test used. Of the two types of tests used in the experiment, the perceptual tests appeared to offer the greater promise for discriminating substantial numbers of such subjects. This implication was evident on the basis of the positive trends observed in the perceptual tests, as contrasted with the negative results among the verbal tests. It might be concluded that the perceptual tests somehow impose conditions on the subject's performance which are more conducive

to tapping the deeper rooted, more basic aspects of personality and behavior as related to the movement response and, accordingly, possess more in common as measures of movement response tendencies.

Of the three perceptual tests, the Rorschach and Behn-Rorschach appeared to enjoy greater mutual relationship in the area of movement production than any other pair. Their mutuality appeared to extend through every movement category, including sum movement—which, incidentally, appeared to be the most stable of the several indices in this relationship, whether observed on an intra- or intertest basis. It may be noted that Epstein, Nelson, and Tanofsky (7), using their extensive series of new inkblots, similarly found sum movement to be the most reliable of the movement indices. The results of this as well as of the present study might be considered to support those authorities who insist that important information may be overlooked in scoring only the human movement responses. The results would suggest, further, that if preference is to be given to any one index, it should perhaps be given to sum movement and not to human movement.

Considering the phi values obtained in the present study, however, in relation to correlation coefficients of comparable grade and in relation to the reliabilities of the two tests, there would appear to be some justification for regarding the relationship between the Rorschach and Behn-Rorschach to be approximately unity in all categories of movement when corrected for attenuation. (Incidental computations, in fact, confirmed this possibility.) There thus appeared to be no theoretical ground for preference between movement categories as far as these

two tests are concerned. It might be concluded that the two tests offer essentially common measures limited in their interrelationship in the several categories of movement only by their variously low reliabilities. While this finding encouraged acceptance of the tests as theoretically parallel forms for observing movement production, it would seem doubtful whether in their present state they can be reliably used as such in individual cases. In stressing this limitation, the results of the experiment essentially agreed with those reported by Eichler (6) and others for the two tests generally.

Similar but somewhat narrower conclusions appeared warranted concerning the Rorschach and Levy Movement Cards on the basis of their significant but low interrelationship in the *M* and *m* categories. It is notable that previous comparisons between these two tests have yielded no significantly positive results. Assessing their own investigations in this respect, King (12), and Allen, Ray, and Poole (1) suggested that the negative findings might have been due to the marked differences in the instructions for the two tests. The results of the present experiment would support such a conclusion and suggest that the two tests do have some mutual validity when less structured directions are substituted for the standard form prescribed for the Levy test.

The contrasting results observed for the two types of tests, perceptual and verbal, lead to some speculation concerning the possible causes. Most readily apparent is the fact that the contrast did not appear to be due to differences in reliability between the two types of tests. The results were, in fact, contrary to what might have been predicted on these grounds, since the verbal tests were on

the whole more reliable than the perceptual tests.

Apparently, the stability of the movement responses as observed in the perceptual tests is sufficiently great as to make itself evident (for some subjects) despite the low reliability of the tests. If so, then the development of more adequate tests of this character, possibly along the lines suggested by Barron (4) might greatly expand their discriminative value. This possibility suggested itself most strongly in connection with the *m* responses which, despite being least reliable, revealed significantly positive results not only with respect to performance in the perceptual tests, but also for the full battery. It is conceivable that a successful effort to raise the reliability of the *m* category could make it the most generally significant; some special study may well be due this category of movement.

As to the possibility of raising the reliability of any of the movement indices, it is interesting that the results obtained by Epstein, Nelson, and Tanofsky (7) discouraged the feasibility of substantially improving reliabilities in the ink-blot tests by lengthening the stimulus card series within practical limits. It might be added that if the results of the present study can be accepted as indicative, then it may not be possible to raise reliabilities enough in tests of this type to permit individual prediction in any case except with reference to specific populations.

Except for the *FM* and *m* categories in Sentence Construction 2, reliabilities in the verbal tests appeared high enough to suggest that the tests might very well measure internally consistent tendencies of some sort. Apparently, however, the conditions imposed on the subject's per-

formance by these tests were not conducive to tapping such deep rooted and stable tendencies as might—except for some few subjects—generalize broadly either among the verbal or from the verbal to the perceptual tests. Or, if basically stable functions were indeed sampled, then the functions must be assumed generally independent of each other and of any sampled by the perceptual techniques.

Dörken (5) suggested that different types of projective test materials or tasks might tap different levels of personality structure. He asserted on this basis that projective techniques might be expected to yield mutually consistent results only when (a) the tests are of the same or of a similar type and hence measure on the same level of personality or (b) when the different levels of personality tapped by different types of tests are themselves consistently organized—a condition he considered likely only for normal subjects.

The general lack of agreement between the verbal and perceptual tests in the present experiment might partially, perhaps, be understood on Dörken's (5) terms. Failure of the majority of the subjects to be consistent within *either* of the two types of tests, however, is not so easily explained. It might even be concluded from this result that, quite contrary to Dörken's (5) principles, the projective techniques may measure on different levels of personality regardless of whether the same or different types of stimulus materials and/or tasks are employed. Some investigators have, in fact, suggested that the projective techniques may sample different levels of personality structure even within the confines of a single test, depending on the subject. Macfarlane and Tuddenham (14) have noted this possibility as, perhaps, a major

problem inherent in all projective techniques and a key question in relation to general theory and practice.

Implications for General Theory and Practice

Among the basic assumptions commonly accepted as logical consequences of the relatively unstructured nature of the projective techniques, Macfarlane and Tuddenham noted the following: (a) "... a belief that the psychological determinants of each and every response are basic and general." (b) "... the belief that projective techniques tap the durable essence of personality equally in different individuals" (14, p. 34). As against these assumptions, the authors noted that nonprojective psychology has long recognized test behavior as partially determined by peripheral and transitory factors. On their own account, they cited unpublished materials suggesting that, while the protocols of some subjects dwelt more on the basic and enduring levels of personality, others dwelt more on superficial levels and sometimes revealed only the protective coverings. In confirming the hypothesis of consistency for some subjects but not for others, the results of the present experiment easily lent themselves to similar interpretation, suggesting that the movement responses in a given projective test might well relate to basic, enduring levels of personality for some subjects but only to peripheral and transitory aspects for others.

The results of the experiment thus joined in suggesting that one of the important consequences of using relatively unstructured techniques may be that the subject is free to determine, consciously or unconsciously, by his own immediate set the level of personality on which he responds and, thereby, to limit the rela-

tion between his performance and basic aspects of personality and behavior.

Such an interpretation, however, involves acceptance of a prior assumption: that the test invariably samples personality functions even when the subject's performance departs from a basic frame of reference. Macfarlane and Tuddenham have expressed it as an explicit principle in "The Assumption of Response Determination." That is, "... that every subject's responses are not the consequence of sheer accident but are determined by psychological attributes of that subject" (14, p. 33). This statement is comparable to that offered by Zubin and Young, namely "... that the test behavior itself is a small sample of life behavior and not a random inexplicable performance" (30, p. 2-5). Macfarlane and Tuddenham (14) stressed the principle as a statement of scientific determinism, which few psychologists would be likely to question. Zubin and Young (30) emphasized it as a postulate which demands proof.

It may be noted that the principle as expressed in both statements predicates the personal relevance of all classes of projective test performance on a hypothesis of nonchance behavior. Treatment of the data in the present study offered a test of the hypothesis and, thereby, of the principle itself, in relation to the movement responses. The results showed that only for a very small number of cases did the movement responses occur with such regularity as to minimize chance occurrence. It might be concluded accordingly that the movement responses might very well be "determined" by chance for some and even the great majority of subjects in certain groups. By the same token, it might seriously be questioned whether the non-

chance hypothesis and, with it, the assumption of response determination can be accepted, without qualification, as generally valid for all classes of projective test behavior.

One of the most important implications of the experiment was thus to stress the need for basic research on the projective techniques with a view toward ultimate verification or revision of their commonly accepted theoretical foundations. The results of the present study suggested that a safer premise for observing projective test behavior might be the proposition that the subject's responses *may or may not* be "determined" by chance. Pending reversal of this implication by further research, it might be concluded that the interpretation of the movement responses and, perhaps, other classes of projective test behavior, as well, may not be justified without establishing in the individual case that the subject's responses actually occur in accordance with nonchance principles. Measures of intra-individual consistency based on multiple or battery testing may thus prove to be indispensable for both experimental and clinical purposes.

SUMMARY

The purpose of the experiment was to examine the assumption that the movement response as observed in projective tests represents something basically stable in the individual's personality and behavior.

It was hypothesized that, if the assumption is tenable, then the individual's comparative level of movement production is consistent over a range of different test situations.

The hypothesis was examined on the basis of the percent human (*M*), animal (*FM*), inanimate (*m*), and sum movement

(*SM*) scores obtained by 82 ostensibly normal male military recruits, 18 to 25 years old, on a battery of three perceptual and three verbal tests: the Rorschach, Behn-Rorschach, Levy Movement Cards, two new Sentence Construction tests, and a modification of Severn's heteronym technique. The data in each category of movement were first examined for evidence of significant interest agreement on whether the subject scored consistently above or below the median on the full battery of tests. Supplementary analyses were made to detect trends and relationships among the three perceptual tests, the three verbal tests, and all possible pairs of tests. Finally, individual differences in variation were compared from one category of movement to another. Two additional hypotheses were examined in this connection: (a) that the degree of variation shown by the subject in one category of movement is consistent with that shown in other categories; (b) that all or a significant majority of subjects are consistent in some one or more categories of movement.

It was concluded from the results that:

1. The conception of movement in all three common categories—*M*, *FM*, and *m*—might well be related to deep rooted, stable characteristics in some but not for the great majority of the subjects.
2. The relationship appeared more likely to be found in tests of the perceptual than of the verbal type, but even on this basis it did not appear likely for all or even a majority of subjects.
3. If the conception of movement bears any significant relationship to psychological processes for more than a minority of subjects, then the relationship is largely specific to certain tests or types of tests and test performance is not generally predictive from one type of test

to another for any individual.

4. Individual differences in variation were not of a generally consistent character from one category of movement to another so as to suggest that they might represent something intrinsically stable in the individual except where the subject is generally consistent in level of movement production.

5. There was no indication in inter-category comparisons that stable individual differences might be generally defined on the basis of selectively consistent tendencies favoring some one or more of the several categories of movement.

The results of the study were considered in relation to others reported in the literature and were cautiously interpreted to suggest that:

1. The conception of movement might

well be relevant for some individuals or groups but not for others; future research might well be directed toward identifying these classes of subjects and providing a basis for discretionary interpretation in clinical practice.

2. Tests of the perceptual type appeared to offer greater promise than those of the verbal type for discriminating substantial numbers of such subjects.

3. The utility of individual differences in variation for diagnostic classification—if warranted, at all—is limited to specific areas of performance and cannot be assumed without establishing the stability of the observed differences.

4. The validity of certain basic assumptions underlying the projective techniques was questioned and their revision suggested.

REFERENCES

1. ALLEN, R. M., RAY, C. D., & POOLE, R. C. The Levy Movement Test: Suggestions for scoring and relationship to the Rorschach movement responses. *J. consult. Psychol.*, 1953, 17, 195-198.
2. ALTUS, W. D. Some correlates of the group Rorschach and the schizophrenia scale of the MMPI among two groups of "normal" college students. *J. consult. Psychol.*, 1948, 12, 375-378.
3. ALTUS, W. D., & THOMPSON, GRACE M. The Rorschach as a measure of intelligence. *J. consult. Psychol.*, 1949, 13, 341-347.
4. BARRON, F. Threshold for perception of human movement in inkblots and its personality correlates. *J. consult. Psychol.*, 1955, 19, 33-38.
5. DÖRKEN, H. JR. Projective tests and the consistency of the personality structure: A pilot study. *J. abnorm. soc. Psychol.*, 1953, 48, 525-531.
6. EICHLER, R. A comparison of the Rorschach and Behn-Rorschach inkblot tests. *J. consult. Psychol.*, 1951, 15, 185-189.
7. EPSTEIN, S., NELSON, JANE V., & TANOFKY, R. Responses to inkblots as measures of individual differences. *J. consult. Psychol.*, 1957, 21, 211-215.
8. HARROWER, MOLLY R., & STEINER, M. E. *Manual for psychodiagnostic inkblots (a series parallel to the Rorschach inkblots)*. Author.
9. HAYS, W., GELLMAN, S., & SLOAN, W. A study of the relationship between the verb-adjective quotient and the Rorschach experience balance. *J. clin. Psychol.*, 1951, 7, 224-227.
10. HUTT, M. L., GIBBY, R. G., MILTON, E. O., & POTTHARST, K. The effect of varied experimental "sets" upon Rorschach test performance. *J. proj. Tech.*, 1950, 14, 181-187.
11. IVES, VIRGINIA, GRANT, MARGUERITE Q., & RANZONI, JANE H. The "neurotic" Rorschachs of normal adolescents. *J. genet. Psychol.*, 1953, 83, 31-61.
12. KING, G. C. Rorschach and Levy Movement responses: a research note. *J. clin. Psychol.*, 1955, 11, 193-195.
13. KLOPPER, B., & KELLEY, D. MC. *The Rorschach technique*. Yonkers: World Book Co., 1946.
14. MACFARLANE, JEAN W., & TUDDENHAM, R. T. Problems in the validation of projective techniques. In H. H. Anderson, & Gladys L. Anderson (Eds.), *An introduction to projective techniques*. New York: Prentice-Hall, 1951. Pp. 26-54.
15. MAYMAN, M. A. A comparative study of the Rorschach, Harrower and Behn-Eschenburg inkblot tests. Unpublished master's thesis, New York Univer., 1947.

16. MELTZOFF, J., & LEVINE, M. The relationship between motor and cognitive inhibition. *J. consult. Psychol.*, 1954, **18**, 355-358.
17. MILLER, D. R. Prediction of behavior by means of the Rorschach test. *J. abnorm. soc. Psychol.*, 1953, **48**, 367-375.
18. MILTON, E. O. The influence of varied experimental sets upon certain Rorschach variables: The human movement variables. Unpublished doctoral thesis. Univ. of Michigan, 1950.
19. RABIN, A. I., & SANDERSON, M. H. An experimental inquiry into some Rorschach procedures. *J. clin. Psychol.*, 1947, **3**, 216-225.
20. RORSCHACH, H. H. *Psychodiagnostics*. Bern, Switzerland: H. Huber, 1942.
21. RUST, R. M. Some correlates of the movement response. *J. Pers.*, 1948, **16**, 369-401.
22. SANFORD, F. H. Individual differences in the mode of verbal expression. Unpublished doctoral dissertation, Harvard College Library, 1941.
23. SEVERN, D. J. Heteronyms for diagnostic use. *J. consult. Psychol.*, 1949, **13**, 144-145.
24. SINGER, J. L., MELTZOFF, J., & GOLDMAN, C. C. Rorschach movement responses following motor inhibition and hyperactivity. *J. consult. Psychol.*, 1952, **16**, 359-364.
25. THORNTON, G. H., & GUILFORD, J. P. The reliability and meaning of Erlebnistypus scores in the Rorschach test. *J. abnorm. soc. Psychol.*, 1936, **31**, 324-330.
26. TUCKER, J. E. Rorschach human and other movement responses in relation to intelligence. *J. consult. Psychol.*, 1950, **14**, 283-285.
27. WALKER, HELEN M., & LEV, J. *Statistical inference*. New York: Holt, 1953.
28. WITTENBORN, J. R. Statistical tests of certain Rorschach assumptions: The internal consistency of scoring categories. *J. consult. Psychol.*, 1950, **14**, 1-19.
29. ZUBIN, J. Symposium on statistics for the clinician. *J. clin. Psychol.*, 1950, **6**, 1-6.
30. ZUBIN, J., & YOUNG, KATHLEEN M. *Manual of projective and cognate techniques*. Madison: College Typing Co., 1948.
31. ZULLIGER, H. *Einführung in den Behn-Rorschach test*. Bern, Switzerland: H. Huber, 1946.

(Accepted for early publication April 12, 1958)

APPENDIX

SENTENCE CONSTRUCTION 2

Stimulus words in order of presentation

1. the	11. then	21. first	31. with
2. all	12. because	22. also	32. in
3. once	13. least	23. any	33. between
4. both	14. or	24. within	34. so
5. most	15. across	25. as	35. another
6. often	16. only	26. though	36. about
7. every	17. although	27. but	37. either
8. a	18. anything	28. behind	38. therefore
9. some	19. always	29. how	39. again
10. each	20. however	30. and	40. very

HETERONYMS⁴ AND MASKING WORDS IN ORDER OF PRESENTATION⁴

Set 1

1. develop	6. <i>conduct</i> *	11. deliver	16. <i>reprint</i>
2. <i>insert</i>	7. examine	12. <i>protest</i>	17. awaken
3. continue	8. <i>increase</i>	13. assemble	18. <i>commune</i> *
4. <i>recruit</i> **	9. accomplish	14. <i>desert</i>	19. resemble
5. discover	10. <i>escort</i> *	15. distinguish	20. <i>offset</i>

Set 2

1. museum	6. <i>suspect</i> *	11. attorney	16. <i>combine</i> *
2. <i>converse</i> *	7. physician	12. <i>convict</i>	17. mechanic
3. apartment	8. <i>extract</i> *	13. horizon	18. <i>discount</i>
4. <i>import</i>	9. cathedral	14. <i>insult</i> *	19. umbrella
5. potato	10. <i>progress</i>	15. piano	20. <i>consort</i>

Set 3

1. together	6. <i>buffet</i>	11. tomorrow	16. <i>conscript</i>
2. <i>perfume</i> *	7. unjustly	12. <i>present</i>	17. wherever
3. extremely	8. <i>survey</i>	13. already	18. <i>ferment</i> *
4. <i>permit</i>	9. whenever	14. <i>compress</i> *	19. precisely
5. beforehand	10. <i>exploit</i>	15. exactly	20. <i>compact</i>

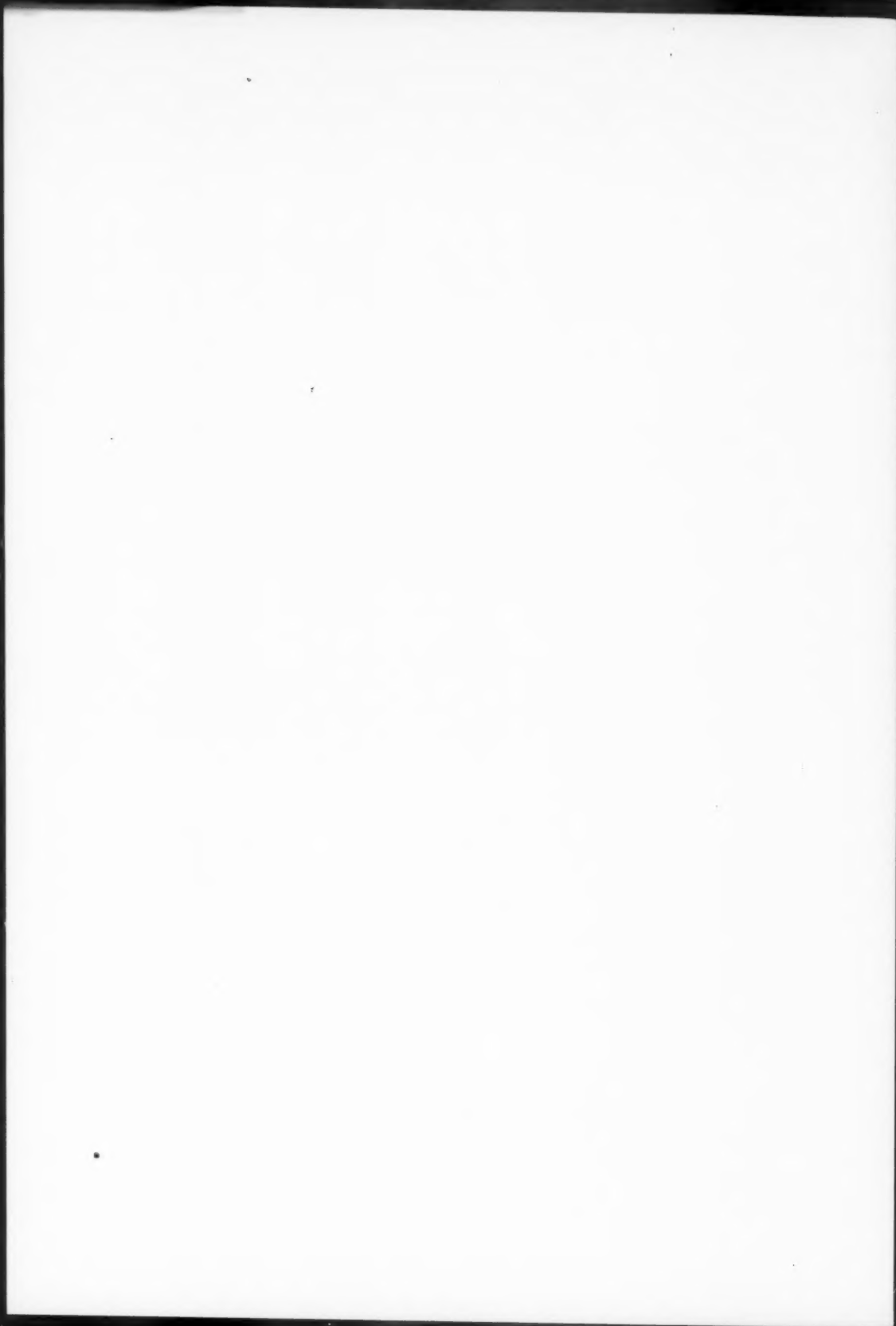
Set 4

1. romantic	6. <i>transfer</i>	11. unhappy	16. <i>console</i>
2. <i>convert</i>	7. dramatic	12. <i>produce</i>	17. abundant
3. familiar	8. <i>project</i> *	13. eternal	18. <i>recount</i>
4. <i>rebel</i>	9. tremendous	14. <i>abstract</i>	19. eleven
5. religious	10. <i>object</i>	15. peculiar	20. <i>digest</i> *

REPRODUCTION OF HETERONYM CARD 2, SET 1

w v n k g i y k g v m j n m o i x d b p l f j v t v
t m l f k c g d z w j l q d r h t a v x v x r z r y
r t s a j k h x f j i n s e r t g h i v w l a z q e
w d d p u v y r u l r g t h h v t c g q f v r f m z
z h w w g r f f b h f s y o g e v x e y n w z k e i
g w x y r p g h c i d y s k h l n v r w u x j k v i
v v i o z x y p t w a f e x a q s q u j f i w s p

⁴ Italicized items: all except those asterisked are from Severn (23). Double asterisked item erroneously included; not a heteronym.



GEORGE BARTLETT COMPANY, INC., MENASHA, WISCONSIN

